PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Sensitivity of rapid antigen tests for COVID-19 during the Omicron variant outbreak among players and staff members of the Japan Professional Football League and clubs: A retrospective observational study
AUTHORS	Murakami, Michio; Sato, Hitoshi; Irie, Tomoko; Kamo, Masashi; Naito, Wataru; Yasutaka, Tetsuo; Imoto, Seiya

VERSION 1 – REVIEW

REVIEWER	Nikolay Mayanskiy Pirogov Medical University
REVIEW RETURNED	15-Sep-2022

GENERAL COMMENTS	In the present paper, Murakami M et al. analyzed sensitivity of rapid antigen tests (RATs) for SARS-CoV-2 against PCR testing in a cohort of 656 football players and staff. They demonstrated an overall RATs sensitivity of 0.63. In the Discussion, the authors speculated that frequent RATs with such sensitivity are more effective than rare PCR testing in several aspects. Apparently, this is a useful conclusion promoting implementation of RATs in various settings. The data comparing RATs and PCR testing are presented in Table 1 (this table should contain %% or actual values for sensitivity and specificity; also, the caption is very confusing: …may be smaller than the actual values …???). Actually, this is the only clear table in the manuscript. Probably, the authors could have limited the results providing only this table reducing their paper to a kind of letter/short communication. The remaining results are confusing. The authors used RATs of different manufacturers as well as various specimen types. Expectedly, in table 2, different kit and different specimen types demonstrated different sensitivity: Abbott 0.73, Roche 0.47, “Either” (what does it mean? Other? Unknown) 0.59; Saliva 0.58, Nasal swab 0.82, “Either or Other” (again, what does it mean?) 0.83. However, in Discussion, the authors stated that “there were no significant differences in the sensitivity irrespective of the manufacturer or sample types including the groups “either” or “either or other”".

REVIEWER	Ellen Hertzmark Harvard University T H Chan School of Public Health, Global Health
REVIEW RETURNED	23-Sep-2022

GENERAL COMMENTS	Review of BMJ Open-2022-067591 This is an interesting paper on a heavily tested population, even though the information on results is unfortunately incomplete.

	How did they compute the confidence intervals for the sensitivity and specificity? I suspect that they used the confidence intervals for single proportions, but they should specify in the Methods section. Arguably, since the denominators are also random variables, it would be better to bootstrap all confidence intervals. Given the relatively small sample, it is not surprising that no predictors of sensitivity met the p<0.05 threshhold. This paper is ready to go after the description of the computation of confidence intervals for the sensitivity and specificity is added to the Methods section.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

>Comments to the Author:
>In the present paper, Murakami M et al. analyzed sensitivity of rapid antigen tests (RATs) for SARS-CoV-2 against PCR testing in a cohort of 656 football players and staff. They demonstrated an overall RATs sensitivity of 0.63. In the Discussion, the authors speculated that frequent RATs with such sensitivity are more effective than rare PCR testing in several aspects. Apparently, this is a useful conclusion promoting implementation of RATs in various settings.

We appreciate your positive and constructive comments. We have revised our manuscript as below.

>The data comparing RATs and PCR testing are presented in Table 1 (this table should contain %% or actual values for sensitivity and specificity; also, the caption is very confusing: …may be smaller than the actual values …???). Actually, this is the only clear table in the manuscript. Probably, the authors could have limited the results providing only this table reducing their paper to a kind of letter/short communication.

To clarify the sensitivity and specificity, we have added the percentages in Table 1. Regarding the caption, we explained the detail in “Participants” section (L113-125 in the revised manuscript with marked changes). We have added the explanation to eliminate the incomprehensibility (L193). We consider that the information in Tables 2 and 3 is important. In particular, as noted in the Introduction and objectives (L72-77; L98-99), whether or not the sensitivity of the rapid antigen tests compared with the PCR test varies with the number of days since the infection has a significant impact on the effectiveness of the test and is essential information for discussing what type of testing regime is effective. Therefore, Tables 2 and 3, and other related descriptions have been retained in the manuscript. In this regard, however, as noted in this comment and the comment below, since some participants were unable to identify the kit manufacturer or PCR sample type, we found it unclear and unhelpful to discuss the relationship between the sensitivity and the kit manufacturer or PCR sample type. We have removed these descriptions (Table 2; L28-29; L40-41; L99-100; L173; L198-200; L228; L237; L269-272).

>The remaining results are confusing. The authors used RATs of different manufacturers as well as various specimen types. Expectedly, in table 2, different kit and different specimen types demonstrated different sensitivity: Abbott 0.73, Roche 0.47, “Either” (what does it mean? Other? Unknown) 0.59; Saliva 0.58, Nasal swab 0.82, “Either or Other” (again, what does it mean?) 0.83. However, in Discussion, the authors stated that “there were no significant differences in the sensitivity irrespective of the manufacturer or sample types including the groups “either” or “either or other”".

“Either” represented the participants whose kit manufacturer or PCR sample types were not identified (L149-154). In the analysis that included options such as “either” or “either or other”, no significant

differences were observed; however, as the reviewer pointed out, the analysis that included the options such as “either” or “either or other” was inappropriate. Since our main objective in this study was to discuss the sensitivity and the duration from the onset of symptoms to testing, we have removed these results and related descriptions (Table 2; L28-29; L40-41; L99-100; L173; L198-200; L228; L237; L269-272).

Reviewer: 2
>Comments to the Author:
>This is an interesting paper on a heavily tested population, even though the information on results is unfortunately incomplete.

We appreciate your positive and constructive comments. We have revised our manuscript as below.

>How did they compute the confidence intervals for the sensitivity and specificity? I suspect that they used the confidence intervals for single proportions, but they should specify in the Methods section. Arguably, since the denominators are also random variables, it would be better to bootstrap all confidence intervals. Given the relatively small sample, it is not surprising that no predictors of sensitivity met the p<0.05 threshhold.
This paper is ready to go after the description of the computation of confidence intervals for the sensitivity and specificity is added to the Methods section.

Following your comment, we have performed the Bootstrap method to estimate the 95%CI of sensitivity and specificity. We have added the explanation in the Methods section (L166-169 in the revised manuscript with marked changes) and corrected the value in Abstract, Results, and Discussion sections (L38; L186; L189-190; L201; L230; L251).

VERSION 2 – REVIEW

REVIEWER	Ellen Hertzmark Harvard University T H Chan School of Public Health, Global Health
REVIEW RETURNED	31-Oct-2022

GENERAL COMMENTS	Review of BMJ Open 2022-076591.R1 I still think that this is a good paper, but that some things need to be clarified. I thank the other reviewer for noting some of them. Specific comments: Word comment Line 46-48 add that those tested were frequently tested. By the way, are the data limited in any way to specific football clubs? (yes, the 23 mentioned) 656 is much lower than the number of people employed in the JFL (over 3600 according to line 100. How many people in the 23 clubs?). I assume that the 656 therefore result from deletions of “not tested on same day” and “both tests negative and not reported to JFL”. I understand that this latter number is unknown, but perhaps it would be useful to note that the total numbers of PCR tests and RATs reported to JFL were n1 and n2, and of these, only 656 were on the same day for a tested person. This last bit belongs in the same section as line 100, and possibly in a figure. The strength here is that you have data from a group of individuals who were tested a lot of times, both routinely and when there were

	symptoms in their organization. 49 This is not a strength. Just a description of what you did. 50-51 To describe this as a strength, change to “We had enough samples to examine the association between the sensitivity of the RAT and the time from the onset of symptoms to testing.” (doesn’t have to be exactly my wording, but the point is that you have a big enough to do it). 52-53 is good. Add o “Not all RATs could be paired with a same-day PCR.” • Many cases in which both tests were negative were not reported to the JFL. • No information on personal characteristics that could be related to test positivity was available. 66-71 The concern is that RATs may be less sensitive early in the disease process (though the authors could phrase this better). Do you mean “infection and testing” or “infection and symptoms”. Obviously testing can’t do anything before it happens. The “contrary to this” does not actually address this issue, but only the relative sensitivity between earlier variants and omicron, and that on a very small sample. 99 how many people are related to the 23 clubs? This might make readers happier that you only have 656 observations. Jan 12 to march 2 is about 7 weeks, so routine testing twice a week should have produced about 14 tests per person. 656 samples could be complete data on at most 47 people. I don’t know whether it is possible to produce a study diagram, describing how many samples there were and why they aren’t in this study (obviously with a big box of neg-neg, number unknown). I think it would make readers happy if 656 was a high fraction of something. 106 Change “have been” to “be” 118 And in infectious disease jargon, testing on days -1 and -2 was for “presymtomatic cases.” 133 I don’t think that the other reviewer will be happy with this wording. I suggest adding “(i.e., sometimes Abbott, sometimes Roche”) after “either”, assuming that this is what the authors mean. 134 Similarly, I suggest adding “(i.e., sometimes saliva, sometimes nasal swab”) after “either”, again assuming that tis is what the authors mean. 157 Is this the first introduction of -1 and -2? If so, perhaps gloss as (i.e., testing a day or 2 before symptom onset). 172 Instead of “See the details,” just repeat the information (because not all samples where both tests were negative were reported to the JFL). Table 1 The title is very long. Probably just “Results of rapid antigen…tests”, but I would like to see “Japanese Football League employees at the beginning of the SARS-Cov-2 omicron outbreak” added to this and all table titles. The sentence stating the results for sensitivity and specificity should be in the text of the results, not the table title. 235 Perhaps add “with isolation of those testing positive”. Testing by itself does nothing if the results are not acted upon. 244-263 This is the real list of limitations that belongs in the bullet points at the beginning of the study. 252 clinical diagnosis—doesn’t this ignore asymptomatic cases? Why would one want to do this? 255-256 Ct values. What would the authors have done if they had known these values? A sensitivity analysis with different Ct cutoffs? 260-263 in what way are these characteristics thought to affect the

	sensitivity of RATs (as opposed to the likelihood of getting Covid-19)?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

>I still think that this is a good paper, but that some things need to be clarified. I thank the other reviewer for noting some of them.

We appreciate your constructive comments. We have revised our manuscript as below. We have asked a native English speaker to correct the texts throughout our revised manuscript again.

Specific comments:

>46-48 add that those tested were frequently tested. By the way, are the data limited in any way to specific football clubs? (yes, the 23 mentioned) 656 is much lower than the number of people employed in the JFL (over 3600 according to line 100. How many people in the 23 clubs?). I assume that the 656 therefore result from deletions of “not tested on same day” and “both tests negative and not reported to JFL”. I understand that this latter number is unknown, but perhaps it would be useful to note that the total numbers of PCR tests and RATs reported to JFL were n1 and n2, and of these, only 656 were on the same day for a tested person. This last bit belongs in the same section as line 100, and possibly in a figure. The strength here is that you have data from a group of individuals who were tested a lot of times, both routinely and when there were symptoms in their organization.

We appreciate your important feedback. “656” represents the number of cases who had PCR and antigen tests performed on the same day. Those cases in which PCR testing was performed but rapid antigen testing was not on the same date, or in which rapid antigen testing was performed but PCR testing was not on the same day, were not included. The number of clubs where these 656 cases were performed was 23 clubs, but we did not limit the number of clubs to collect the cases, but as a result, this study included the cases among 23 clubs, for which PCR and antigen tests were performed on the same date.

We have the information on the number of rapid antigen tests conducted by the J-League for regular tests, but we do not obtain the information on the number of rapid antigen tests conducted voluntarily by each club. We also do not have the information on the cases in which PCR testing was performed but rapid antigen testing was not on the same date.

Following the reviewer’s suggestion, we have added the explanation in Strength and Limitation with a consideration of possible number of bullet points (no more than 5 bullet points) (P3L46-49 in the revised manuscript with marked changes). We have also added these explanations and created a figure that shows these numbers in the main text (P7L108-122; Figure 1). We have also noted the number of players and staff members in the 23 clubs in the main text (P7L121-122).

>49 This is not a strength. Just a description of what you did.

We have deleted this sentence.

>50-51 To describe this as a strength, change to “We had enough samples to examine the association between the sensitivity of the RAT and the time from the onset of symptoms to testing.” (doesn’t have to be exactly my wording, but the point is that you have a big enough to do it).

We have revised this sentence (P3L54-55).

>52-53 is good.

Thank you for your comment.

>Add

> “Not all RATs could be paired with a same-day PCR.”

> Many cases in which both tests were negative were not reported to the JFL.

> No information on personal characteristics that could be related to test positivity was available.

Thank you for your suggestions. We need to consider the possible number of bullet points regarding Strengths and Limitations (no more than 5 bullet points). We have therefore added sentences regarding the first and third points (P3L56-58). In this revising process, we have deleted the sentence about the fifth bullet point: “Since the participants were professional sport players and staff members, cautions are required in applying the findings of this study in general or other populations.”

Regarding the second point (“Many cases in which both tests were negative were not reported to the JFL”), we would like to add that perhaps there was a misunderstanding. As explained above and in the explanations added to the main text (P7L108-122, Figure 1), this survey does not cover cases in which antigen and PCR tests were not performed on the same date. In some of the rapid antigen tests performed, the PCR test was performed on the same date: we included cases when both tests were performed on the same date, although it is true that some cases in which both tests on the same date were negative may not have been reported to the JFL.

>66-71 The concern is that RATs may be less sensitive early in the disease process (though the authors could phrase this better). Do you mean “infection and testing” or “infection and symptoms”. Obviously testing can’t do anything before it happens. The “contrary to this” does not actually address this issue, but only the relative sensitivity between earlier variants and omicron, and that on a very small sample.

We appreciate your comments. Here, we would like to mention as follows:

1. high-frequency testing using rapid antigen testing is expected as a testing system, because it can identify infected presymptomatic individuals with a high viral load from the time of infection until the onset of symptoms; 2. it has been pointed out that the sensitivity of the rapid antigen test (especially during the first few days after infection) may be reduced for Omicron variants; 3. there is concern that antigen testing may be less effective as a routine testing system during an outbreak of Omicron variants; 4. however, findings regarding the decreased sensitivity of rapid antigen testing in omicron variants have been inconsistent.

We have revised our manuscript to explain these points (P5L69-80).

>99 how many people are related to the 23 clubs? This might make readers happier that you only have 656 observations. Jan 12 to march 2 is about 7 weeks, so routine testing twice a week should have produced about 14 tests per person. 656 samples could be complete data on at most 47 people. I don’t know whether it is possible to produce a study diagram, describing how many samples there were and why they aren’t in this study (obviously with a big box of neg-neg, number unknown). I think it would make readers happy if 656 was a high fraction of something.

We appreciate you for pointing this out. As noted above, we have added the information about the number of testing and the number of players and staff in the 23 clubs (P7L108-122; Figure 1). Furthermore, we consider that the number of “unique participants” covered in this study is important. Therefore, we prepared a table showing the date and number of cases per club covered in this study (Table 1). The same person was never subjected to antigen tests or PCR tests more than once on the same day: the number of cases performed in a given club on a given day represents the number of unique participants (no duplicates). Also, the same person did not belong to different clubs. Therefore, the maximum number of cases done on given days in each club represents the minimum possible number of unique participants in the club and the sum of the minimum possible number of unique participants in clubs (n = 309) represents the minimum possible number of participants in this study. We have added the explanations in the main text (P8L127-133; Table 1).

>106 Change “have been” to “be”

We have revised this part (P8L126).

>118 And in infectious disease jargon, testing on days -1 and -2 was for “presymtomatic cases.”

We have added the explanation (P10L151-152).

133 I don’t think that the other reviewer will be happy with this wording. I suggest adding “(i.e., sometimes Abbott, sometimes Roche”) after “either”, assuming that this is what the authors mean.

We have revised our manuscript following your kind suggestion (P11L168).

>134 Similarly, I suggest adding “(i.e., sometimes saliva, sometimes nasal swab”) after “either”, again assuming that tis is what the authors mean.

We have revised our manuscript (P11L169).

>157 Is this the first introduction of -1 and -2? If so, perhaps gloss as (i.e., testing a day or 2 before symptom onset).

As described above, we have added the explanation(P10L151-152).

>172 Instead of “See the details,” just repeat the information (because not all samples where both tests were negative were reported to the JFL).

We have revised our manuscript (P13L210-211).

>Table 1 The title is very long. Probably just “Results of rapid antigen…tests”, but I would like to see “Japanese Football League employees at the beginning of the SARS-Cov-2 omicron outbreak” added

to this and all table titles. The sentence stating the results for sensitivity and specificity should be in the text of the results, not the table title.

We have revised the title in all figure table captions according to your suggestion (Figure 1; Tables 1-4). We have deleted the sentence stating the results for sensitivity and specificity, which was already described in the result section (P12L200-202).

>235 Perhaps add “with isolation of those testing positive”. Testing by itself does nothing if the results are not acted upon.

We have added this point (P17L276).

>244-263 This is the real list of limitations that belongs in the bullet points at the beginning of the study.

We have added the limitations regarding two points in your comments above (P18L287-290). The limitation regarding the third point in your comments above (“No information on personal characteristics that could be related to test positivity was available”) was described (P18L301-303). As described above, since the number of bullet point in Strength and Limitations should be less than or equal to five, we could not list all the limitations in the Strengths and Limitations section.

>252 clinical diagnosis—doesn’t this ignore asymptomatic cases? Why would one want to do this?

We mentioned this limitation, because it has been reported that the sensitivity of the PCR test against clinical diagnosis is not always 100% even among the symptomatic cases. We have added the explanation of this point (P18L299).

>255-256 Ct values. What would the authors have done if they had known these values? A sensitivity analysis with different Ct cutoffs?

We have added the explanation by citing a refence (P19L304-306).

>260-263 in what way are these characteristics thought to affect the sensitivity of RATs (as opposed to the likelihood of getting Covid-19)?

We agree with you. Here, we have modified the statement to a possible difference in familiarity with the test because of the group's training on the test, rather than a health population bias (P19L310-314).

VERSION 3 – REVIEW

REVIEWER	Ellen Hertzmark Harvard University T H Chan School of Public Health, Global Health
REVIEW RETURNED	16-Dec-2022

GENERAL COMMENTS	Review of BMJopen-2022-067591.R2 The authors have done almost everything I wanted them to do. I especially appreciate the extra tables and the new figure. My only comment now is that with at least 309 distinct subjects and only 656 test pairs, it is unlikely that there are large clusters. Further, it is unlikely that the correlation between test results (or sensitivity of tests) has substantial within-person correlation. I am happy to say that this paper is now ready for publication.